Criteria for NTP Immunotoxicology Studies

Report from the Immunotoxicology Criteria Working Group of the National Toxicology Program Board of Scientific Counselors

Submitted by:
Nancy Kerkvliet, Ph.D., chair
Mitzi Nagarkatti, Ph.D., rapporteur
Michael Woolhiser, Ph.D., rapporteur

The National Toxicology Program (NTP) Board of Scientific Counselors (BSC) is a federally chartered external advisory group that provides input on the scientific merit of NTP's programs and activities. The Immunotoxicology Criteria Working Group (ICWG) of the NTP BSC was formed in May 2008. The purpose of the ICWG was to evaluate the utility of having specific criteria for describing the results from individual NTP immunotoxicology studies to indicate the strength of the evidence for their conclusions. The ICWG was composed of 13 scientists representing academia, industry, and government. Dr. Nancy Kerkvliet, Oregon State University, a member of the NTP BSC, chaired the ICWG. Drs. Michael Luster, National Institute for Occupational Safety and Health; Kimber White, Virginia Commonwealth University; Susan Elmore, NTP; Dori Germolec, NTP; and Paul Foster, NTP served as advisors to the ICWG. Drs. Mitzi Nagarkatti, Michael Woolhiser, and Lori White, NTP Executive Secretary, served as rapporteurs. Also in attendance at the meeting was Dr. Mary Wolfe, NTP Federal Official. The ICWG roster is attached [Appendix A]. The ICWG met August 13 and 14, 2008 at the Crystal City Marriott, 1999 Jefferson Davis Highway, Crystal City, VA.

The NTP developed draft criteria for describing results of NTP studies that were modeled after the NTP criteria used to evaluate toxicology and carcinogenicity studies. Drs. Dori Germolec and Paul Foster from the Toxicology Branch of the NTP are the lead scientists for this effort. Prior to the ICWG meeting, the draft criteria were evaluated internally by NTP staff.

Dr. Foster opened the ICWG meeting by providing the background for the development of criteria by NTP. Dr. Germolec presented information regarding NTP immunotoxicology testing strategies and a discussion of the draft immunotoxicology criteria. Materials provided to the ICWG included: the draft criteria [Attachment B], a set of case studies for testing the utility and applicability of the draft criteria for reaching conclusions on NTP immunotoxicology studies, a list of issues for discussion by the ICWG [Attachment C], and the carcinogenicity criteria [Attachment D]. The ICWG was given the following charge:

> *Evaluate the suitability and utility of the proposed criteria for describing the results from individual NTP immunotoxicology studies to indicate the strength of the evidence for their conclusions.*

The case studies that were reviewed by the committee members had been provided by both NTP and members of the ICWG and were purposely designed to reflect the type of data sets that have been used to draw conclusions about immunotoxicity. Many were designed to have significant data gaps that would lower the strength of the evidence for immunotoxicity. The ICWG members worked separately on the case studies and scored them using the draft criteria. The results were tallied so that the group could view the degree of concordance (or lack thereof). The ensuing discussions revealed the thought process behind each member's score, and proved quite constructive in refining the criteria so that the boundaries between categories were as clear as reasonably possible. Much of the discussion was intended to refine the specific wording of the

criteria and was driven by the case studies The outcome of these deliberations was the following revised draft criteria.

Levels of Evidence for Evaluating Immune System Toxicity

Clear Evidence of Toxicity to the Immune System

- Is demonstrated by data that indicate a clear treatment-related (considering the magnitude and the dose-response) effect on more than one functional parameter and/or a disease resistance assay that is not a secondary effect of overt systemic toxicity, or
- Is demonstrated by data that indicate treatment-related effects on one functional assay and additional endpoints that indicate biological plausibility.

Some Evidence of Toxicity to the Immune System

- Is demonstrated by data that indicate a treatment-related effect on one functional parameter with no other supporting data, or
- Is demonstrated by data that indicate treatment-related changes in multiple non-functional parameters without robust changes in a functional immune parameter or a disease resistance assay, or
- Is demonstrated by data that indicate non-dose-related effects on functional parameters or a disease resistance assay with other data providing biological plausibility.

Equivocal Evidence of Toxicity to the Immune System

- Is demonstrated by data that indicate non-dose-related effects on functional parameters or a disease resistance assay without other data providing biological plausibility, or
- Is demonstrated by data that indicate treatment-related changes in a single non-functional parameter without changes in a functional immune parameter or a disease resistance assay, or
- Is demonstrated by data that indicate immune effects at dose(s) that produce evidence of overt systemic toxicity, or
- Is demonstrated by data that are conflicting in repeat studies.

No Evidence of Toxicity to the Immune System

- Is demonstrated by data from studies with appropriate experimental design and conduct that indicate no evidence of biologically relevant changes in immune parameters.

The ICWG determined that the "inadequate study" level included in the draft criteria would not be considered because NTP would not conduct studies with qualitative or quantitative limitations that could not be interpreted for immunotoxicity. The ICWG held further deliberations and developed points of discussion to be considered by NTP in applying their final criteria.

Points for NTP consideration in applying the criteria

- Immunotoxicity is defined in the context that immune responses can be enhanced or suppressed by toxicants. As such, treatment-related effects consistent with immunosuppression and immunostimulation will be considered in hazard identification.

- The characterization of immunotoxicity must consider the impact of overt toxicity (e.g., effects on the immune system are not the direct effects of chemical treatment, but are indirect effects mediated via stress and/or other treatment-related responses).

- The characterization of immunotoxicity must consider the intended pharmacology of the chemical. Immunotoxicity is reserved for unintended immunosuppression or immunostimulation.

- It is recognized that recovery may occur following cessation of treatment. However, even transient immune effects that may be observed during treatment or shortly thereafter are important for hazard identification.

- Biological plausibility for immunotoxicity must be considered in the context of the nature of the response, the magnitude of the response, and the pattern of the response, as well as the current understanding of immune system structure and function.

- Functional changes in an immune response should usually be weighted more heavily than non-functional changes.

- Based on historical experience, *in vivo* assays are more sensitive in detecting immunotoxicity than *in vitro* assays. *In vivo* assays also take into account the metabolism of the toxicant that may either reduce or increase immunotoxicity.

- Results in one species or one sex are considered sufficient for evidence of immunotoxicity.

- The purpose of the criteria is for hazard identification only, not risk assessment.

ICWG Comments

Dr. Kerkvliet will present a draft of the ICWG meeting report to the NTP BSC at its meeting in November 2008. The meeting was a great success for many reasons. The composition of the committee included academia, industry, and government experts with expertise in immunotoxicology, immunotoxicity testing and regulatory needs. All of the committee members were fully engaged in the process, actively participating in the discussions. Following deliberations on the case studies, the working group reviewed the draft criteria and systematically revised them based on applications of the criteria to the case studies. The draft criteria prepared by NTP staff served as a framework for the discussions and the final criteria produced at the meeting. Examples used in the case studies were submitted by ICWG participants and NTP staff, and were vital to illustrate how the criteria could be successfully applied to experimental data. Each participant was asked to work independently, to read each case study, and to assign a draft criteria level based on the data available and then NTP staff collated the responses. Each case was then discussed and various committee members shared their rationale for choosing a particular criteria level. The earnest discussions that ensued allowed for revisions of the draft criteria and a final product that was acceptable to all members of the committee.

Appendix A

NTP Board of Scientific Counselors Immunotoxicology Criteria Working Group

Working Group Members

Robert Benson, Ph.D.
Drinking Water
US Environmental Protection Agency
1595 Wynkoop
Denver, CO 80202-1129

Scott Burchiel, Ph.D.
University of New Mexico
College of Pharmacology and Toxicology
1 UNM MSC09 5360
Albequerque, NM 87131

Jeanine Bussiere, Ph.D, DABT
Dept of Toxicology
Amgen, Inc.
One Amgen Center Drive MS 29 2 A
Thousand Oaks, CA 91320

Vicki L. Dellarco, Ph.D.
Office of Pesticide Programs
US Environmental Protection Agency
1200 Pennsylvania Ave. NW, 7510P
Washington, DC 20460

Rodney Dietert, Ph.D.
Department of Microbiology
& Immunotoxicology
Cornell University
C5 135 VMC
College of Veterinary Medicine
Ithaca, NY 14853

Michael Holsapple, Ph.D., ATS
ILSI Health and Environmental Sciences Institute
One Thomas Circle
Washington, DC 20005

Robert V. House, MSPH, Ph.D.
Science and Technology
DynPort Vaccine Company LLC
64 Thomas Johnson Drive
Frederick, MD 21702

Nancy Kerkvliet, MS, D.Phil. (chair)
Oregon State University
2750 SW Campus Way, Room 1007
Corvallis, OR 97331

L. Peyton Myers, Ph.D.
Division of Antiviral Products
US Food and Drug Administration
10903 New Hampshire Avenue
WO 22 RM 6373
Silver Spring, MD 20903

Mitzi Nagarkatti, Ph.D.
Professor & Chair
Department of Pathology, Microbiology
and Immunology
University of South Carolina
School of Medicine
6439 Garners Ferry Road
Columbia, SC 29208

Peter Thomas, Ph.D.
Covance, Inc.
3301 Kinsman Blvd.
Madison, WI 53704

Michael R. Woolhiser, Ph.D.
Toxicology & Environmental Research
& Consulting
Dow Chemical Company
Building 1803 Washington Street
Midland, MI 48674

Yung Yang, Ph.D.
Health Effects Division
Office of Pesticide Program
US Environmental Protection Agency
1200 Pennsylvania Ave, NW, (MC 7509P)
Washington, DC 20460

Advisors to the Working Group

Michael I. Luster, Ph.D.
Senior Advisor
National Institute for Occupational Safety & Health
1095 Willowdale Rd.
Morgantown, WV 26505

Kimber White, Ph.D.
Department of Pharmacology and Toxicology
Virginia Commonwealth University
PO Box 980613
527 N. 12th Street
Richmond, VA 23298

Susan Elmore, DVM, MS, DACVP
Cellular and Molecular Pathology Branch
National Institute of Environmental Health Sciences
National Institutes of Health
111 T.W. Alexander Drive, PO Box 12233
Research Triangle Park, NC 27709

Dori Germolec, Ph.D.
Immunology Discipline Leader
National Institute of Environmental Health Sciences
National Institutes of Health
111 T.W. Alexander Drive, PO Box 12233
Research Triangle Park, NC 27709

Paul Foster, Ph.D.
Toxicology Branch Chief
National Institute of Environmental Health Sciences
National Institutes of Health
111 T.W. Alexander Drive, PO Box 12233
Research Triangle Park, NC 27709

NTP Executive Secretary

Lori White, Ph.D., PMP
Office of Liaison, Policy, and Review
National Institute of Environmental Health Sciences
National Institutes of Health
111 T.W. Alexander Drive, PO Box 12233
Research Triangle Park, NC 27709

NTP Federal Official

Mary Wolfe, Ph.D.
Director, Office of Liaison, Policy, and Review
National Institute of Environmental Health Sciences
National Institutes of Health
111 T.W. Alexander Drive, PO Box 12233
Research Triangle Park, NC 27709

Appendix B

Draft Criteria for Evaluating Immunotoxicity Data from NTP Studies

1. **Clear Evidence of Toxicity to the Immune System**. Is demonstrated by the results of a study or studies in one or more species that indicate a clear treatment-related effect on disease resistance assays and/ or other sensitive immune parameters (e.g. a decrease in antigen-specific antibody production) that is not a secondary effect of overt systemic toxicity. Effects in multiple endpoints that suggest biological plausibility (i.e. alterations in NK cell cytotoxicity and NK cell numbers) would also provide clear evidence of immune toxicity.

2. **Some Evidence of Toxicity to the Immune System**. Is demonstrated by a study or studies indicating a treatment-related change in immune parameter(s) in which the spectrum of the response is less than that required for strong evidence. For example, there may be statistically significant changes in the histology of the immune tissues or leukocytes counts without any clear effects on associated functional parameters (e.g. no changes in lymphocyte subpopulations), changes in molecular or genomic endpoints suggestive of altered immune function or where deficits have been noted in one or more less sensitive end points (e.g. lymphoproliferative responses) with no associated histological changes in the spleen, thymus, bone marrow or lymph nodes.

3. **Equivocal Evidence of Toxicity to the Immune System** is demonstrated by a study or studies that are interpreted as showing marginal deficits in immune parameters that may be treatment-related (e.g. statistically significant changes in one or more parameters at middle or low doses in the absence of other supportive data).

4. **No Evidence of Toxicity to the Immune System** is demonstrated by a study or studies with appropriate experimental design and conduct that are interpreted as showing no evidence of chemically related deficits in immune parameters.

5. **Inadequate study of immunotoxicity** is demonstrated by a study that because of major qualitative or quantitative limitations cannot be interpreted as valid for showing the presence or absence of immunotoxicity.

Other key points to be noted in the assessment

- The evidence of deficits in immune parameters may be supported by other data from *in vivo* animal studies (e.g. 14, 90-day studies) which may show gross and histopathological effects in the primary or secondary immune tissues of a nature and extent indicating a high likelihood of an adverse effect on immune function (e.g. thymic atrophy). Chemical relationship with other known immune system toxicants would also add to the evidence supporting a "clear evidence" description.
- Occurrence of common versus rare immune deficits
- Use of historical control data to place concurrent control into perspective and estimate population background incidence of immune parameters.
- Similarity between clinical endpoints in humans and measures evaluated in rodent studies.

Appendix C

Issues for Discussions by NTP BSC Immunotoxicology Criteria Working Group

1. Conclusions statements for NTP studies are hazard-based, and not risk-based, to facilitate comparison across chemicals using the same study types. These conclusion statements are voted upon by the NTP Board of Scientific Counselors in its advisory role to the NTP Executive Committee, which contains representatives from our sister regulatory agencies that can use this information in quantitative risk assessment decisions.
2. It would be helpful if we could model conclusion criteria for non-cancer studies based on that currently employed for the NTP carcinogenicity studies (attached), to generate some consistency in approach and wording for both the Board and the public.
3. NTP staff recognizes that for many of the non-cancer toxicity studies, we are dealing with multiple (interrelated) endpoints very different from cancer studies. Thus, the NTP cancer study approach to levels of evidence in drawing study conclusions will require some "finessing" to achieve the desired level of consistency.
4. NTP staff also recognizes the desirability to use a graded (hazard identification) conclusion scheme, such that a single positive finding does not necessarily result in the highest level of conclusion. We have considered those endpoints that affect overall function to merit the highest level of conclusion (clear evidence of toxicity). So, there maybe a statistically significant, dose-related decrease in some endpoint (for example, $CD4^+$ leukocyte counts in an immunotoxicology study), but without a concomitant effect on animal function (e.g., disease resistance or functional immune parameters), it would not merit the clear evidence category.

Appendix D

Definition of Carcinogenicity Results

The National Toxicology Program describes the results of individual experiments on a chemical agent and notes the strength of evidence for conclusions regarding each study. Negative results, in which the study animals do not have a greater incidence of neoplasia than control animals, do not necessarily mean that a chemical is not a carcinogen, inasmuch as the experiments are conducted under a limited set of conditions. Positive results demonstrate that a chemical is carcinogenic for laboratory animals under the conditions of the study and indicate that exposure to the chemical has the potential for hazard to humans. Five categories of evidence of carcinogenic activity are used in the Technical Report series to summarize the strength of the evidence observed in each experiment: two categories for positive results ("Clear Evidence" and "Some Evidence"); one category for uncertain findings ("Equivocal Evidence"); one category for no observable effects ("No Evidence"); and one category for experiments that because of major flaws cannot be evaluated ("Inadequate Study"). These categories of interpretative conclusions were first adopted in June 1983 and then revised in March 1986 for use in the Technical Reports series to incorporate more specifically the concept of actual weight of evidence of carcinogenic activity. For each separate experiment (male rats, female rats, male mice, female mice), one of the following quintet is selected to describe the findings. The categories refer to the strength of the experimental evidence and not to either potency or mechanism. In 1987, Haseman et al., (Environ. Health Perspec. 74: 229-235, 1987) reclassified earlier studies (Technical Reports No. 2-200, 202-205) according to the new five category system. We have appended this information to the end of those numbered abstracts enabling a uniformity across the entire collection of technical reports.

Clear Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing a dose-related (i) increase of malignant neoplasms, (ii) increase of a combination of malignant and benign neoplasms, or (iii) marked increase of benign neoplasms if there is an indication from this or other studies of the ability of such tumors to progress to malignancy.

Some Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing a chemical-related increased incidence of neoplasms (malignant, benign, or combined) in which the strength of the response is less than that required for clear evidence.

Equivocal Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing a marginal increase of neoplasms that may be chemically related.

No Evidence of Carcinogenic Activity is demonstrated by studies that are interpreted as showing no chemical-related increases in malignant or benign neoplasms.

Inadequate Study of Carcinogenic Activity is demonstrated by studies that because of major qualitative or quantitative limitations cannot be interpreted as valid for showing either the presence or absence of carcinogenic activity.

When a conclusion statement for a particular experiment is selected, consideration must be given to key factors that would extend the actual boundary of an individual category of evidence. This should allow for incorporation of scientific experience and current understanding of long-term carcinogenesis studies in laboratory animals, especially for those evaluations that may be on the borderline between two adjacent levels. These considerations should include:

* The adequacy of the experimental design and conduct;
* Occurrence of common versus uncommon neoplasia;
* Progression (or lack thereof) from benign to malignant neoplasia as well as from preneoplastic to neoplastic lesions;
* Some benign neoplasms have the capacity to regress but others (of the same morphologic type) progress. At present, it is impossible to identify the difference. Therefore, where progression is known to be

a possibility, the most prudent course is to assume that benign neoplasms of those types have the potential to become malignant;

* Combining benign and malignant tumor incidences known or thought to represent stages of progression in the same organ or tissue;
* Latency in tumor induction;
* Multiplicity in site -specific neoplasia;
* Metastases;
* Supporting information from proliferative lesions (hyperplasia) in the same site of neoplasia or in other experiments (same lesion in another sex or species);
* The presence or absence of dose relationships;
* The statistical significance of the observed tumor increase;
* The concurrent control tumor incidence as well as the historical control rate and variability for a specific neoplasm;
* Survival-adjusted analyses and false positive or false negative concerns;
* Structure-activity correlations; and
* In some cases, genetic toxicology.

Earlier designations include:
P = Positive; E = Equivocal; N = Negative;